



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/647,203	08/21/2003	Alexander Franz	16113-1230001 / GP-038-00	1475
26192	7590	11/12/2008	EXAMINER	
FISH & RICHARDSON P.C. PO BOX 1022 MINNEAPOLIS, MN 55440-1022			SHAH, PARAS D	
			ART UNIT	PAPER NUMBER
			2626	
			NOTIFICATION DATE	DELIVERY MODE
			11/12/2008	ELECTRONIC

**Please find below and/or attached an Office communication concerning this application or proceeding.**

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

PATDOCTC@fr.com

### Office Action Summary

**Application No.**

10/647,203

**Applicant(s)**

FRANZ ET AL.

**Examiner**

PARAS SHAH

**Art Unit**

2626

**Period for Reply** -- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

- 1) ☒ Responsive to communication(s) filed on 09/26/2008.
- 2a) ☒ This action is **FINAL**. 2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

- 4) ☒ Claim(s) 55-69 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 55, 56, 58-61, 63-66, 68 and 69 is/are rejected.
- 7) ☒ Claim(s) 57, 62 and 67 is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

**Application Papers**

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on \_\_\_\_\_ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.  
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

**Priority under 35 U.S.C. § 119**

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some \* c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
  2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
  3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

**Attachment(s)**

- 1) ☐ Notice of References Cited (PTO-892)
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) ☐ Information Disclosure Statement(s) (PTO/SF/08)  
Paper No(s)/Mail Date \_\_\_\_\_
- 4) ☐ Interview Summary (PTO-413)  
Paper No(s)/Mail Date \_\_\_\_\_
- 5) ☐ Notice of Informal Patent Application
- 6) ☐ Other: \_\_\_\_\_

### **DETAILED ACTION**

1. This communication is in response to the Argument and Amendments filed on 09/26/2008. Claims 55-69 remain pending and have been examined, with claims 1-54 being cancelled. The Applicants' amendment and remarks have been carefully considered but they do not place the application in condition for allowance.
2. All previous objections and rejections directed to the Applicant's disclosure and claims not discussed in this Office Action have been withdrawn by the Examiner.

### ***Response to Amendments and Arguments***

3. Applicant's arguments (pages 7-8) filed on 09/26/2008 with regard to claims 55-69 have been fully considered and are moot in view of new grounds for rejection.

### ***Claim Objections***

4. Claims 60-64 are objected to because of the following informalities: The limitation of "operable to cause data" should be changed to "causing a" since the former does not yield a positive recitation. Appropriate correction is required.

### ***Claim Rejections - 35 USC § 112***

5. The following is a quotation of the first paragraph of 35 U.S.C. 112:

The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention.

6. Claims 60-64 are rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the written description requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention. The claimed limitation of "computer program product" is not defined in the Specification. The term computer program is the only limitation found in the Specification.

7. The following is a quotation of the second paragraph of 35 U.S.C. 112:

The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

8. Claims 60-64 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention. It is unclear as to what the Applicant is seeking to claim with the claim limitation of a "computer program product". It is unclear as to whether the computer program product contains the computer readable medium or just program code, if the former, then how is the computer readable medium encoded on the computer readable medium. Hence, for the purposes of compact prosecution it was interpreted to mean the computer program.

### ***Claim Rejections - 35 USC § 103***

9. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically taught or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the

prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

10. Claims 55, 59, 60, 64, 65, and 69 are rejected under 35 U.S.C. 103(a) as being unpatentable over Su *et al.* (In *Proceedings of the 32nd Annual Meeting on Association For Computational Linguistics* 1994) in view of Frantzi *et al.* ("Extracting Nested Collocations").

As to claims 55, 60, and 65, Su *et al.* teaches a computer-implemented method for identifying compounds in text, comprising:

extracting a vocabulary (see page 244, 2<sup>nd</sup> full paragraph, sect. Simulation, (1<sup>st</sup> paragraph), line 5-8, compound list) of tokens(see page 244, Table 1) from text(see page 243, left column, 2<sup>nd</sup> paragraph, line 6);

identifying a plurality of unique n-grams in the text (see page 245, right column, "Simulation," 1st paragraph, compound list is modified or rebuild after a new compound word is detected. The compounds having plurality of lengths is obvious in document being studied.), each n-gram being an occurrence in the text of n sequential tokens, each token being found in the vocabulary (see page 244, left column, lines 4-18, relative frequency of the n-gram is computed. It is obvious to one skilled in the art that the n-gram is associated with the respective frequency);

dividing each n-gram into n-1 pairs of two adjacent segments, where each segment consists of at least one token (see page 243, right column, 2<sup>nd</sup> full paragraph, Mutual Information, where words are used to determine word association measure.);

for each n-gram, calculating a likelihood of collocation for each pair of segments of the n-gram (page 243, right column, equation, specifically, the probabilities in the formula that are calculated, numerator and denominator) and determining a score (page 243, right column, line 8, value obtained, lambda) for the n-gram based on a lowest calculated likelihood of collocation (see page 243, right column, line 23, lambda has a lowest value of 0) ;

identifying a set of n-grams having scores above a threshold (see page 243, right column, line 23, lambda has a lowest value of 0, value of lambda greater than zero identifies a compound word)); and

adding the identified set of n-grams as compound tokens to the vocabulary (see page 245, right column, 2<sup>nd</sup> paragraph, line 7, compound list) and removing constituent tokens that occur in the added compound tokens from the vocabulary (see page 244, left column, Relative Frequency Count paragraph, relative frequency is used to prevent when multiple occurrences occur).

However, Su *et al.* does not specifically teach the use of iterating from  $n > 2$  down to  $n = 2$  where  $n$  decreases by one each iteration and in each iteration performing the actions. It should be noted that Su *et al.* does suggest using window sizes of two or three for n-gram determination (see page 243, left column, 1<sup>st</sup> paragraph).

Frantzi *et al.* does teach the use iterating from  $n > 2$  down to  $n = 2$  where  $n$  decreases by one each iteration and in each iteration performing the actions (page 43, right column, "The algorithm ...", 2<sup>nd</sup> full paragraph, code underneath

and page 44, entire left column-right column, numbered item 5) (e.g. From the cited reference it is seen that the n-gram starts from some maximum limit and then proceeds to a lower order n-gram. The n-gram is decremented and takes into account the frequency of occurrence in order to determine a candidate collocation by the determination of a C value.)

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the finding of compounds words in a corpus as taught by Su *et al.* with the backward iteration as taught by Frantzi *et al.* The motivation to have combined the references involves the ability to systematically determine the likelihood of collocation and extract the unextracted collocations that occur (see Abstract) and thus making the process automatic.

As to claims 61 and 66, it would have been obvious to one of ordinary skilled in the art to have implemented the method of claim 55 into a computer readable medium whereby a processor executes the program from the medium. Further, Su suggests the use of a computing system (see page 245, right column, Simulation) and so does Frantzi (see page 43, right column, code on left hand side).

As to claims 59, 64, and 69 Su *et al.* in view of Frantzi *et al.* teaches all of the limitations as in claims 55, 61, and 66, above.

Furthermore, Su teaches where identifying a plurality of unique n-grams in the text comprises skipping n-grams appearing in a list of known compounds

(see page 244, left column, Relative Frequency Count paragraph, relative frequency is used to prevent when multiple occurrences occur, thus skipping n-grams appearing in list ).

11. Claims 56, 58, 61, 62, 66 and 68 are rejected under 35 U.S.C. 103(a) as being unpatentable over Su *et al.* in view of Frantzi *et al.* as applied to claims 1, 6, 13, and 24 above, and further in view of Manning (The MIT Press 1999).

As to claims 56, 61, and 66 Su *et al.* in view of Frantzi *et al.* teaches all of the limitations as in claims 1, 6, 13, and 24 above.

Furthermore, Su *et al.*, teaches where the likelihood ratio  $\lambda$  is computed by:  $\lambda = (P(x|M_c) * P(M_c)) / (P(x|M_{nc}) * P(M_{nc}))$  (see Su *et al.*, page 243, right column, line 9 (equation)) (e.g. It should be noted that the reference uses a different notation, but the same result and definitions are used, where the numerator is the n-gram produced by a compound result and the denominator is the result produced by a non-compound result. The formula can be changed to account for various distributions (Gaussian, Binomial).

However, Su *et al.* in view of Frantzi *et al.* do not specifically teach the likelihood ratio given by  $\lambda = L(H_i) / L(H_c)$ .

Manning shows the use of the likelihood ratio (see equation 5.10) (e.g. The equation is given in log form. The logs can be omitted to obtain the desired formula. The numerator is the independent hypothesis and the denominator is the dependence hypothesis.)



It would have been obvious to one of ordinary skilled in the art to have modified finding of compounds in a text corpus as taught by Su *et al.* and Frantzi *et al.* with the formula as taught by Manning. The motivation to modify the former is for collocation discovery (see Manning, page 172, sect. 5.3.4, 3<sup>rd</sup> paragraph, lines 1-4).

As to claims 58, 63, and 68, Su *et al.* in view of Frantzi *et al.* teaches all of the limitations as claim 13, 16, and 17 above.

Su *et al.* in view of Frantzi teach a system for identifying compounds through measure of association.

However, Su *et al.* in view of Frantzi do not specifically teach the representation of the independence and collocation hypothesis.

Manning does teach the explanations of these two types of hypothesis (see page 172, sect. 5.3.4, bullet items) (e.g. It should be noted that the independence hypothesis is given by hypothesis 1 and the dependence or collocation hypothesis by hypothesis 2. The  $w_2$  and  $w_1$  can be interpreted as the tokens since the reference deals with a text corpus).

It would have been obvious to one of ordinary skilled in the art to have modified the finding of compound words in a text corpus as taught by Su *et al.* and Frantzi *et al.* with the inclusion of the two hypothesis as taught by Manning. The motivation to modify the former is for collocation discovery (see Manning, page 172, sect. 5.3.4, 3<sup>rd</sup> paragraph, lines 1-4). Further, the use of the formula

presented by Manning would require an explanation of frequency for each type of hypothesis in order to find the likelihood ratio (definition of likelihood ratio).

### ***Allowable Subject Matter***

12. Claims 57, 62, and 67 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

13. The following is a statement of reasons for the indication of allowable subject matter: none of the prior art references alone or in combination teaches or fairly suggests the limitations where the limitations "L(H<sub>c</sub>) is computed ... in accordance with

the formula:  $\arg \max_{L(H_c)} \frac{L(t_1, t_2, \text{formcompound})}{L(n - \text{gramdoesnotformcompound})}$ " as seen in claims 19 and 30.

### ***Conclusion***

14. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of

Art Unit: 2626

the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

15. Any inquiry concerning this communication or earlier communications from the examiner should be directed to PARAS SHAH whose telephone number is (571)270-1650. The examiner can normally be reached on MON.-THURS. 7:00a.m.-4:00p.m. EST.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Patrick Edouard can be reached on (571)272-7603. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/P. S./  
Examiner, Art Unit 2626

11/06/2008

Application/Control Number: 10/647,203

Page 11

Art Unit: 2626

/Patrick N. Edouard/

Supervisory Patent Examiner, Art Unit 2626